DOI: 10.1079/WPS2006125

Molecular markers for the assessment of chicken biodiversity

J. HILLEL[1]*, Z. GRANEVITZE[1], T. TWITO[2], D. BEN-AVRAHAM[1], S. BLUM[1], U. LAVI[2], L. DAVID[3], M.W. FELDMAN[4], H. CHENG[5] and S. WEIGEND[6]

[1]The Hebrew University of Jerusalem, Faculty of Agriculture, Rehovot 76100, Israel; [2]ARO-Volcani Center, P.O.B.6 Bet Dagan 50250 Israel; [3]Stanford Genome Technology Center, Stanford University, California, 94304, USA; [4]Department of Biological Sciences, Stanford University, California, 94305, USA; [5]USDA-ARS, Avian Disease and Oncology Laboratory, 3606 E. Mount Hope Rd. East Lansing, MI 48823, USA; [6]Institute for Animal Breeding Mariensee, Federal Agricultural Research Centre (FAL); 31535 Neustadt; Germany
***Corresponding author: hillel@agri.huji.ac.il**

Three main aspects of chicken biodiversity are dealt within this report: (a) cluster analysis based on autosomal microsatellites, (b) microsatellites on the sex chromosomes, and (c) SNP-based biodiversity.
(a) Cluster analysis of autosomal microsatellites: **We used 29 microsatellites to genotype 2000 chickens randomly selected from 65 different populations representing various chicken types and various geographical regions. The computer program Structure placed the 65 populations into clusters that are in agreement with their geographic origin and breed history. Only at two predefined clusters, there is little admixture between non-commercial populations originating from Asia and those from Europe. In contrast, commercial broilers and brown egg layers appeared as admixed populations of these two main gene pools. Increasing the number of clusters resulted in generation of specific clusters of commercial lines, having very low admixture with other clusters. In addition, we identified seven mixed populations, each of which shared portions of their genome with several other genetic clusters.**
(b) Microsatellites on the sex chromosomes: **We predicted 173 potential microsatellites on chromosome W by in-silico analysis of the chicken genome assembly (version WASHUC1). Twenty five microsatellites of the highest sequence quality were tested in the lab for gender specificity. Unexpectedly, PCR products were generated in both sexes. Moreover, 14 selected microsatellites were mapped (using the East Lansing reference panel) and in all cases, the "W specific" microsatellites were mapped to chromosome Z and except for one locus, to the same ~6 cM region. We conclude that the draft assembly for chromosome W is quite inaccurate.**

This article is based on data presented at the 4th European Poultry Genetics Symposium, 6-8 October 2005, Dubrovnik, Croatia

World's Poultry Science Journal, Vol. 63, March 2007
Received for publication May 8, 2006
Accepted for publication June 5, 2006

(c) SNP-based biodiversity: **Ten distinct chicken breeds were genotyped at 145 single nucleotide polymorphisms (SNPs) located at 14 random DNA fragments and twenty five, each from different and unlinked genes. Microsatellite genotypes of the same ten breeds were used for comparison. Applying bootstrap values as the criterion for tree's reliability, we found that: (1) increasing the number of SNPs had a higher impact on the reliability of the analysis than increasing the number of individuals per population, and (2) the bootstrap values of phylogenetic un-rooted trees based on microsatellites were relatively low.**

Keywords: biodiversity; bioinformatics; chicken; microsatellites; phylogeny; poultry; SNP; W-chromosome

Introduction

DNA variation between genomes of individuals, families, populations and breeds represents the genetic diversity within a given farm animal species. Improvement of our insight into the mechanisms underlying genetic diversity may assist in the reconstruction of domestication events, assessment of genetic relationships between populations, and of genetic variation within populations. In addition, it may identify important DNA sequences for breeding purposes. Such insights can be gained by genotyping samples from the worldwide chicken gene pool. Moreover, such studies are required for the development of a strategy for optimal sampling of the genetic variation to be conserved.

Chicken biodiversity

Within the order of *Galliformes*, the jungle fowl is classified into four species: red jungle fowl (*G. gallus*), gray jungle fowl (*G. sonnerati*), green jungle fowl (*G. varius*) and Ceylon jungle fowl (*G. lafayettei*) (Delacour, 1977; Haward and Moore, 1984; Johnsgard, 1986). The red jungle fowl, which is believed to be the progenitor of the domesticated chicken, has its widest distribution in east Asia, from Pakistan through China, Eastern India, Burma, most of Indo-China, and on the islands of Sumatra, Java and Bali (Crawford, 1990). Existing poultry varieties comprise a wide range of breeds and strains that have evolved in the process of domestication and breeding. Since domestication, chickens have been distributed to various countries, continents and cultures. Breeding of poultry for commercial purposes using highly efficient selection programmes has resulted in a few highly specialized lines dominating today's world market. In turn, the genetic diversity of chickens has been restricted to a few specialized commercial breeds and a vast range of non-commercial chicken breeds.

DNA-based markers provide powerful tools for estimating genetic relatedness between and within animal populations. Various types of DNA markers have been used, including RFLP, RAPD, minisatellites and microsatellites (Plotsky *et al.*, 1995; Vanhala *et al.*, 1998; Zhou and Lamont, 1999; Romanov and Weigend, 2001; Sharma *et al.*, 2001). These studies are important for understanding the history of species, and for testing hypotheses regarding evolutionary processes (Vishwanathan *et al.*, 2004; Gaines *et al.*, 2005).

Differences in the extent of polymorphism at different regions of the genome may reflect different evolutionary histories. Hence, information regarding various genomic regions, assessed by different types of genetic markers, may increase our understanding of different evolutionary forces.

SNP-based phylogeny

SNPs are single base changes in the DNA sequence, with the rare allele having a frequency of 1% or greater. In practice, most SNPs are bi-allelic, due to the low frequency of single nucleotide mutations. These are estimated in mammals to occur a rate of $1x10^{-9}$ and $5x10^{-9}$ per nucleotide per generation (Martinez-Arias *et al.*, 2001).

Based on the available draft of the chicken genome sequence, about 3.3 million SNPs have been identified by comparing three domestic individuals (a broiler, a layer and a Chinese silkie) and an individual of the wild red jungle fowl. The mean nucleotide diversity was found to be about 1 per 200 bp when comparison was made between red jungle fowl and domestic lines, as well as among and within domestic lines (International Chicken Polymorphism Map Consortium. 2004). It is desirable to estimate the reliability of these SNPs and to sift candidate SNPs for the assessment of genetic variation and evaluation of relationships within and among chicken populations.

W chromosome

Avian females are the heterogametic gender carrying both the Z and the W chromosomes while males are the homogametic gender (ZZ). The non-recombining portions of chromosome W exist only in females and are maternally inherited. As such, W-specific sequences are valuable for the phylogenetic analysis of avian species, as has already been demonstrated for the human genome by studies based on the paternally inherited chromosome Y (Shen *et al.*, 2004).

It is assumed that both Z and W chromosomes evolved from a common autosomal ancestor (Ellegren and Carmichael, 2001). In the process of evolution, some regions of chromosomes Z and W became inaccessible to recombination, probably due to chromosomal inversions (Handley *et al.*, 2004). Identification of W-specific regions is important for other purposes, including the reconstruction of the sex chromosomes, their evolution, and gender identification.

Three main aspects of chicken biodiversity are dealt with:

1) Clustering of breeds based on autosomal microsatellite data generated in a broad worldwide chicken gene pool.
2) Bioinformatics-based search for microsatellites on the W chromosome.
3) Phylogenetic analysis based on SNPs located within gene regions, SNPs found in random non-gene DNA fragments and microsatellites using varying number of loci and individuals.

Materials and methods

CHICKEN POPULATIONS

Sixty-five chicken populations representing a wide range of populations originating from various continents have been subject to various selections and management regimes (*Table 1*). The individuals (samples from twenty chicken populations), and marker information (22 loci) are listed in a biodiversity database established by the European collaboration AVIANDIV[1] project (http://w3.tzv.fal.de/aviandiv/; Weigend *et al.*, 1998). The set of chicken breeds jointly analysed in this study was further complemented by new

[1]AVIANDIV EC Contract No. BIO4-CT98-0342 (1998-2000); Weigend, S (Coordinator), M.A.M. Groenen, M. Tixier-Boichard, A. Vignal, J. Hillel, K. Wimmers, T. Burke, and A. Mäki-Tanila

DNA samples including 28 fancy chicken breeds of different phylogenetic origin in Germany, three layer lines, twelve Chinese native breeds, one population from Vietnam, and another one from Malawi. We aimed to sample the same number of males and females for a total of about thirty chickens per population. Regarding the fancy breeds collected in Germany, blood samples were collected, on average, from seven flocks per breed kept by different breeders. Samples of Chinese native breeds were randomly taken from conservation flocks kept at the Poultry Institute, Academy of Chinese Agricultural Sciences, Yangzhou, P. R. China. Chickens were randomly selected from farms in Malawi communal areas, and from three villages in Maison District of Sonla Province in Northwest Vietnam. In summary, a total, of 65 chicken populations comprised of one subspecies of the red jungle fowl, *Gallus gallus gallus*, 52 non-commercial chicken populations of various origins and breeding management, seven commercial pure bred layer lines, four broiler lines, and one inbred line. Overall, 2,000 birds were analysed.

GENOTYPING

DNA samples were obtained either from the AVIANDIV DNA bank or from fresh blood samples, using standard DNA isolation procedures. Prior to genotyping, the concentration of the DNA was standardized to 20 ng/µl in TE solution.

Twenty nine microsatellites distributed as uniformly as possible throughout the chicken genome, were genotyped individually by PCR. The markers and their genomic position are listed in *Table 2*. Genotyping the 20 AVIANDIV populations at some of the 29 markers was performed on ABI sequencers at several labs of the AVIANDIV project. Genotyping of the remaining 45 populations, was carried out using a semi-automated sequencer LICOR (LICOR Biotechnology Division, Lincoln, Nebraska, USA). For this analysis, one of the two primers was labelled with either IRD700 or IRD800 (MWG-Biotech, Ebersberg, Germany). Allele-size was assessed by the RFLPscan package (Scanalytics, Division of CSP, Billerica, MA). Allele scoring was standardized by samples of known alleles at each locus.

CLUSTER ANALYSES

The clustering algorithm implemented in the *Structure* package (Pritchard *et al.*, 2000) was used to analyze population structure of the gene pool under study. The underlying model assumes that individuals can be assigned to a cluster on the basis of the multilocus genotype of the individual and the estimated allele frequencies of clusters, without any prior information regarding ancestry. The number of clusters (K), has to be defined by the user *a priori,* and structures are inferred through several runs with varying values for K. The proportion of each individual's genome having ancestry in each cluster (membership coefficient), is estimated by the admixture algorithm. Since the approach is not deterministic, several repeated runs may provide different cluster solutions. Therefore, we ran one hundred repeats (100) and chose the most frequent solutions (*Figure 1*) by calculating the matrix of similarity coefficients using *SimCoeff* software (Rosenberg *et al.*, 2002) among all repeats for a given K-value. In order to determine the number of iterations and the burn-in period needed for each solution, we created a curve of likelihood *vs.* iterations and burn-in steps (iterations and burn-in steps varied between 10,000 and 2.5 million steps). The curves were sinusoidal rather than saturated. We therefore used 50,000 iterations after 20,000 burn-in steps in all the analyses.

Structure was applied to the genotypes of the 29 markers for the 2,000 chickens from 65 different populations. The number of clusters (K) varied between 2 and 6, with 100 runs for each *K* value. We calculated the matrix of similarity coefficients - *C*. A similarity threshold was set at 0.95. We considered two replicates to be identical when *C* between them was above the threshold. The most frequent solution among the 100 repeats was chosen as the most suitable solution.

SNP GENOTYPING

We randomly chose 50 SNPs within gene regions from the chicken SNP database, which is based on ESTs alignment (http://chicksnps.afs.udel.edu/). The SNPs validity was tested by Mass-Array (Sequenom Inc. San Diego, CA, USA). Genotyping of several (three to five) individuals from each of the 20 populations (80 individuals altogether) revealed that only 58% (29) of these SNPs, were truly polymorphic in our tested populations. Moreover, we sequenced fragments from 58 genes taken from NCBI database (http://www.ncbi.nlm.nih.gov) and EST databases (http://www.chickest.udel.edu/; http://www-.chick.umist.ac.uk/). Alignment of these sequences revealed 92 SNPs within 30 genes (28 genes don't contain SNPs).

Out of these SNPs described above, each located in a different clone, we chose 25 polymorphic unlinked SNPs (the most reliable ones; *Table 3*), and genotyped them on a total of 150 individuals (five or 10 individuals from each of the 20 populations mentioned above).

In addition, we analyzed the genotype data of 145 SNPs obtained by sequencing 14 random DNA fragments of birds from the same 10 populations. This information was generated in the European AVIANDIV project (see details above).

PHYLOGENETIC CLADOGRAMS

Genetic distances between breeds were calculated from the proportion of shared alleles (PSA) using *Microsat* (http://hpgl.stanford.edu/projects/microsat). Bootstrap values were calculated based on 100-1000 repeats; since no differences were obtained between 100 and 1000 repeats, we performed the entire analyses on 100 repeats. Unrooted cladograms were obtained by *neighbour-joining* (based on 100 genetic distance matrices), using the *neighbour* program from *PHYLIP* package (Felsenstein, 1992). A consensus tree was obtained using the *consensus* program from *PHYLIP* package.

IDENTIFICATION OF MICROSATELLITES

The 5Mb W chromosome sequence from NCBI (WASHUC1) was searched for microsatellites by the program *Tandem Repeats Finder* (Benson, 1999). A set of 173 Simple Sequence Repeats (SSRs), were chosen on the basis of having a repeat unit size of 2-6 nucleotides and a repeat count of more than seven. Selected loci were compared with the *chicken* genomic sequence database (NCBI), by BLAST (Altschul *et al.*, 1990) for novelty and for probable W specificity.

GENDER SPECIFICITY TEST FOR MICROSATELLITE LOCI

Twenty five microsatellites were tested for gender specificity. Primers were designed by *PRIMER 3* (Rosen and Skaletsky, 2000). PCR was carried out on at least four males and four females. The PCR comprised of: 35 cycles of: 94° C for 1 min., 54-58°C (depending on the primer) for 1 min, 72°C for 1 min. The products were electrophoresed on a 1% agarose gel.

MAPPING

Primers were designed to amplify specific microsatellites. Markers that were informative in the East Lansing reference panel were mapped by multipoint analysis of 52 progeny genotypes, and more than 1200 previously mapped loci.

Results and discussion

CHICKEN GENE POOL

We have studied the biodiversity of 2000 chickens from 65 populations that were

collected from around the globe and genotyped at 29 autosomal microsatellites. The genomic composition of chicken populations was characterized using the clustering algorithm of *Structure* and its accompanying software. Most of the 65 populations clustered according to their geographic origin and breed history. The order of the evolutionary divergence of the six clusters generated by *Structure* is presented schematically in *Figure 1*. For K = 3 (three clusters), the populations originating from Asia, Europe, and the commercial lines (Broilers and Brown egg Layers) were clearly separated. This indicates that commercial chicken lines are distinct from the resource populations which may carry genetic features that have not yet been applied in current breeding programs. Clustering of genetic resource populations might provide the basis for assessment of evolution and genetic diversity. Furthermore, it may facilitate the search for specific traits (such as resistance to various diseases), between genetically distinct groups of populations. Assessment of genetic structure of a wide gene pool may therefore provide useful information for evaluating chicken genetic resources (Weigend and Romanov, 2001).

For K = 6 (six clusters), we identified seven populations, which shared parts of their genome with relatively many clusters (*Table 1*, cluster 7). Three of these populations, have previously been found to be highly polymorphic (Hillel *et al.*, 2003).The Red Jungle Fowl may have been the progenitor of the domesticated chicken while others may reflect genetic admixture from several origins. *Figure 1* demonstrates a possible order of the populations' splits, which generated the six main clusters detected by *Structure*.

MICROSATELLITES ON THE SEX CHROMOSOMES

We chose 25 "supposedly W-specific"-microsatellite sequences to be tested for gender specificity. Unexpectedly, PCR products were generated with DNA from both genders. Moreover, 14 of them were mapped to chromosome Z, and except for one microsatellite, to the same ~6 cM region (Ben-Avraham *et al.*, 2006).

The conclusion is that the draft assembly of chromosome W contains errors. The finding that the majority of our mapped microsatellites were located within the five large supercontigs that span from 721,718 bp to 4,842,841 bp of chromosome W, suggests that the majority of this chromosome is incorrectly placed. This result corroborates other data suggesting the existence of inaccuracies in the current genome assembly, particularly on the sex chromosomes (Ben-Avraham *et al.*, 2006).

FACTORS AFFECTING THE ASSESSMENT OF CHICKEN BIODIVERSITY

The effect of the DNA markers (Microsatellites and SNPs), the number of individuals and the number of loci, on the reliability of the phylogenetic trees, was tested by seven cladograms (*Table 4*). The repeatability of the cladograms and their capability to distinguish between populations were assessed by the average bootstrap value (BV) of each cladogram. A value of 60 was set as a threshold for significance of the repeatability. The following results are based on pairwise comparisons between some of these cladograms:

1. 1 *vs.* 2 – Considering 60 as a threshold for significance, twenty nine microsatellites are not sufficient for the assessment of 10 populations, even when the number of individuals per population was increased from 10 (mean bootstrap value [BV] = 47) to 30 (BV = 42).
2. 2 *vs.* 3 –145 SNPs distributed along 14 DNA fragments result in a more repeatable cladogram (BV = 76) than that based on 29 microsatellites (BV = 47); The number of individuals per population in both cases was 10.
3. 3 *vs.* 4 – Much better resolution is obtained when the number of SNPs is increased

from 14 (BV = 29) to 145 (BV = 76) although they were all located on the same 14 DNA fragments.

4. 5 *vs.* 6 - Analysis based on 25 SNPs (each from a different gene), is not affected by reducing the number of birds from 10 (BV = 73) to 5 (BV = 70). The two cladograms are very similar in shape and in bootstrap values.
5. 5 *vs.* 7 – Analysis based on 10 birds per population is highly affected by reducing the number of markers from 25 (BV = 73) to 12 (BV = 43).

Based on the above detailed comparisons, the following conclusions can be drawn:

a. Twenty-nine microsatellite markers are not sufficient to build a reliable un-rooted phylogenetic tree.
b. Increasing the number of SNPs from 1 to about 10 per DNA fragment does improve repeatability of the cladogram.
c. Unrooted trees based on 10 populations represented by 10 individuals per population, have similar repeatability whether generated by 25 SNPs in 25 genes, or by 145 SNPs in 14 random DNA fragments.
d. In these ranges of individuals and markers, increasing the number of SNPs has a better impact on the reliability of the trees than increasing the number of individuals per population.

Concluding remarks

Assessment of the chicken biodiversity becomes feasible and important with the availability of the chicken sequence. The reliable assessment of population clusters, and the availability of SNPs affecting the variation of agriculturally important traits, would provide tools to identify DNA sequences important for breeding.

Our conclusions are: (1) Population clustering based on microsatellites and appropriate clustering algorithms (such as *Structure)* is reliable. (2) Phylogeny assessment based on SNPs should rely on a large number of marker loci with a moderate number of individuals per population. (3) The current low quality of the chicken genome sequence, does not yet allow reliable phylogeny assessment, based on female-specific sequences.

Acknowledgments

We wish to thank the following colleagues who provided the blood samples. Without their help this study could not have been performed: Dr. G. Baldane, Padova, Italy; Dr. J. Baumgartner, Research Institute, Ivanka pri Dunaji, Slovakia; Dr. Y. Bondarenko, Research Institute; Borky, Kharkiv, Alaska; Dr. J.L. Campo, INIA, Madrid, Spain; D. Cassuto, ANAK, Israel; Dr. K. Cywa-Benko, Research Institute, Balice n Cracow, Poland; Dr. E. Eythorsdottir, RALA, Israel; Y. Jego, Hubbard-ISA, France; Dr. J. Kathle, NLH, Norway; Dr. I. Moiseyeva, Research Institute, Gubkin, Moscow, Russia; Dr. G. Pidone, Calatami, Italy; Prof. R. Preisinger Lohmann, Germany; M. Protais, Hubbard-ISA, France; P. Rault, SYSAAF, Tours, France; Mrs. V. Roberts, UK; Dr. P. Sorensen, DIAS, Denmark; Dr. I. Szalay, Hungary; Prof. Therdchai Vearasilp, Thailand; Dr. B. Thorp Ross Breeders, Aviagene, UK; Dr. P. Trevil, Biopharmacy Institute, Prague, Czech Republic; F. van Sambeek, Hendrix, Ospel, The Netherlands; Dr.Vilhelmova, Czech Republic; J. Vincent, Cobb, UK; Dr. G. Virag, Research Institute, Godollo, Hungary; Armin Six, Germany, German Poultry Fancy Breeding Organisation; Chen Guohong, Yangzhou University, P. R. China; Ngo Thi Kim Cuc, NIAH Hanoi, Vietnam; Clemens B. A. Wollny

Univ. Göttingen, Germany; Farai Muchadeyi and Timothy Gondwe, Univ. Harare, Zimbabwe

References

ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. and LIPMAN, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

BEN-AVRAHAM D., BLUM, S., GRANEVITZE, Z., WEIGEND, S., CHENG, H. and HILLEL, J. (2006) W-specific microsatellite loci detected by in silico analysis map to chromosome Z of the chicken genome. *Animal Genetics* **37**:180-181.

BENSON, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acid Research* **27**(2): 573-580.

CRAWFORD, R.D. (1990) Origin and history of poultry species. *In:* Poultry Breeding and Genetics. R.D. Crawford Ed. Elsevier, New York, NY. p1-41.

DELACOUR, J. (1977) The Pheasants of the world, 2nd edition. Spur Publication, Hindhead, Surrey.UK.

ELLEGREN, H. and CARMICHAEL, A. (2001) Multiple and independent cessation of recombination between avian sex chromosomes. *Genetics* **158(1)**: 325-331.

FELSENSTEIN, J. (1992) Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. *Genetical Research* **60(3)**: 209-220.

GAINES, C.A., HARE, M.P., BECK, S.E. and ROSENBAUM, H.C. (2005) Nuclear markers confirm taxonomic status and relationships among highly endangered and closely related right whale species. *Proceedings. Biological Science*s **272**: 533-542.

HANDLEY, L.J., CEPLITIS, H. and ELLEGREN, H. (2004) Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167(1)**: 367-376.

HAWARD, R. and MOORE, A. (1984) A complete checklist of birds of the world, revised edition. Macmillian, London, UK.

HILLEL, J., GROENEN, M.A., TIXIER-BOICHARD, M., KOROL, A.B., DAVID, L., KIRZHNER, V.M., BURKE, T., BARRE-DIRIE, A., CROOIJMANS, R.P., ELO, K., FELDMAN, M.W., FREIDLIN, P.J., MAKI-TANILA, A., OORTWIJN, M., THOMSON, P., VIGNAL, A., WIMMERS, K. and WEIGEND, S. (2003) Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. *Genetics, Selection, Evolution* **35**: 533-557.

INTERNATIONAL CHICKEN POLYMORPHISM MAP CONSORTIUM (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**: 717-722.

JOHNSGARD, P.A. (1986) The Pheasants of the world. Oxford University Press, New York USA.

MARTINEZ-ARIAS, R., MATEU, E., BERTRANPETIT, J. and CALAFELL, F. (2001) Profiles of accepted mutation: from neutrality in a pseudogene to disease-causing mutation on its homologous gene. *Human Genetics* **109(1)**: 7-10.

PLOTSKY, Y., KAISER, M.G. and LAMONT, S.J. (1995) Genetic characterization of highly inbred chicken lines by two DNA methods: DNA fingerprinting and polymerase chain reaction using arbitrary primers. *Animal Genetics* **26**: 163-170.

PRITCHARD, J.K., STEPHENS M. and DONNELLY P.J. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

ROMANOV, M.N. and WEIGEND, S. (2001) Analysis of genetic relationships between various populations of domestic and jungle fowl using microsatellite markers. *Poultry Science* **80**: 1057-1063.

ROZEN, S. and SKALETSKY, H. (2000) Primer3 on the WWW for general users and for biologist programmers. pages *In:* Krawetz S, and Misener S. eds. Bioinformatics Methods and Protocols: Methods in Molecular Biology p.365-386. Humana Press, NJ.

ROSENBERG, N.A., PRITCHARD, J.K., WEBER, J.L., CANN, H.M., KIDD, K.K., ZHIVOTOVSKY, L.A. and FELDMAN, M.W. (2002) Genetic structure of human populations. *Science* **298**: 2381–2385

SHARMA, D., APPA RAO, K.B., SINGH, R.V. and TOTEY, S.M. (2001) Genetic diversity among chicken breeds estimated through randomly amplified polymorphic DNA. *Animal Biotechnology* **12**: 111-120.

SHEN, P., LAVI, T., KIVISILD, T., CHOU, V., SENGUN, D., GEFEL, D., SHPIRER, I., WOOLF, E., HILLEL, J., FELDMAN, M. and OEFNER, P.J. (2004) Reconstruction of patri- and matri-lineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Human Mutation* **24**: 248-260.

VANHALA, T., TUISKULA-HAAVISTO, M., ELO, K., VILKKI, J. and MAKI-TANILA, A. (1998) Evaluation of genetic variability and genetic distances between eight chicken lines using microsatellite markers. *Poultry Science* **77**: 783-790.

VISHWANATHAN, H., DEEPA, E., CORDAUX, R., STONEKING, M., USHA RANI, M.V. and MAJUMDER, P.P. (2004) Genetic structure and affinities among tribal populations of southern India: a study of 24 autosomal DNA markers. *Annals of Human Genetics* **68**: 128-138.

WEIGEND, S., GROENEN, M.A.M., TIXIER-BOICHARD, M. VIGNAL, A., HILLEL, J., WIMMERS, K., BURKE, T. and MÄKI-TANILA, A. (1998) Development of Strategy and Application of Molecular Tools to Assess Biodiversity in Chicken Genetic Resources. AVIANDIV; EC Contract No. BIO4-CT98-0342.

WEIGEND, S. and ROMANOV. M.N. (2001). Current strategies for the assessment and evaluation of genetic diversity in chicken resources. *World's Poultry Science Journal* **57**: 275-287.

ZHOU, H. and LAMONT, S.J. (1999) Genetic characterization of biodiversity in highly inbred chicken lines by microsatellite markers. *Animal Genetics* **30**: 256-264.

Note from the author

In the chicken genome sequence release of May 2006 (WASHUC2.1), many of the errors reported in the section MICROSATELLITES ON THE SEX CHROMOSOMES in this article, have been corrected.

Table 1 Chicken populations (65) and clusters (6).

Population Name	Cluster[1]	Population Name	Cluster	Population Name	Cluster
AB line, high*	1	Luyuan	3	Rheinlaender	4
Brown egg layer C*	1	Tibetan	3	Schlotterkaemme	4
Brown egg layer D*	1	Wugu	3	Westf. Totleger	4
Brown egg layer A	1	Xianju	3	Vorwerk	4
Rhodelaender	1	Xiaoshan	3	C line*	5
Broiler dam line D*	2	You	3	Fayoumi*	5
Broiler sire line B*	2	H'mong chicens	3	Green legged Partidge*	5
Tr. Naked Neck*	2	Bergische Kraeher	4	Bedouin*	6
Asil	2	Brabanter	4	Iceland landrace*	6
Broiler dam line A	2	Brakel	4	Jaerhoens*	6
Broiler sire line A	2	Deutsche Sperber	4	Line Sarcoma Suscept.*	6
Marans	2	Friesenhuhn	4	Padova*	6
New Hampshire	2	Hamburger Lackh.	4	Sc. Ref. Population*	6
Sundheimer	2	Hamburger Sprenkel	4	White egg layer A*	6
Brahma	3	Italiener Triesdorf	4	Gallus gallus gallus*	7
Cochin	3	Italiener rebh.	4	Godollo Nhx*	7
Baier	3	Italiener schw.	4	Orlov*	7
Chahua	3	Krueper	4	Kastilianer	7
Dagu	3	Lakenfelder	4	Malaien	7
Dou	3	Ost. Moewen	4	Thuer. Barthuehner	7
Gushi	3	Paduaner	4	Malawi	7
Langshan	3	Ramelsloher	4		

* Chicken populations collected during the AVIANDIV project

[1]*Clusters* are based on microsatellite typing (see Material and Methods section for details): **1** - Brown Egg Layers, **2** – predominantly Broilers, **3** – native Chinese breeds or breeds with recent Chinese origin, **4** - predominantly breeds of North-West European derivation, **5** - predominantly breeds of Middle East background, **6** - predominantly breeds of Primarily Mediterranean background, and **7** – multi-cluster (breeds share several clusters)

Table 2 Description of the 29 microsatellites.

Microsatellite	Chromosome	Map Position [cM]	GenBank Accession	Multiplex Set
MCW0248	1	19	G32016	X1
MCW0111	1	118	L48909	X3
ADL0268	1	288	G01688	X1
MCW0020	1	460	L40055	X5
LEI0234	2	50	Z94837	X3
MCW0206	2	104	AF030579	X7
MCW0034	2	233	L43674	X2
MCW0222	3	85	G31997	X2
MCW0103	3	201	G31956	X7
MCW0016	3	247	L40041	X3
LEI0166	3	300	X85531	X1
MCW0037	3	317	L43676	X3
MCW0295	4	75	G32051	X3
LEI0094	4	153	X83246	X1
MCW0098	4	217	L40074	X6
MCW0078	5	93	L43686	X6
MCW0081	5	151	L43636	X2
MCW0014	6	50	L40040	X4
MCW0183	7	86	G31974	X4
ADL0278	8	94	G01698	X1
MCW0067	10	59	G31945	X6
ADL0112	10	120	G01725	X4
MCW0216	13	47	AF030586	X1
MCW0104	13	74	L43640	X5
MCW0123	14	45	L43645	X5
MCW0080	15	49	G54425	X5
MCW0330	17	41	G32085	X6
MCW0165	23	1	L43663	X5
MCW0069	26	47	L43684	X2

Multiplexes:
X1 ADL0278, ADL0268, LEI0094, MCW0248, MCW0216;
X2 MCW0081, MCW0034, MCW0069, MCW0222, MCW0295;
X3 MCW0111, MCW0037, MCW0016, LEI0166, LEI0234;
X4 MCW0183, ADL0112, MCW0014;
X5 MCW0165, MCW0020, MCW0123, MCW0104; MCW0080
X6 MCW0078, MCW0067, MCW0330, MCW0098;
X7 MCW0206, MCW0103;

Table 3 Description of 25 SNPs within gene regions.

#	Locus ID	SNP	SNP location and Amino acid change[1]	Physical map[2]		SNP type[3]	NCBI GI / Clone ID[4]	Gene name / Best blast hit[5]
				Position[a]	Chrom. [b]			
1	211	G/A	5' flanking region	47995260	+3		1401326	Gallus gallus oestrogen receptor gene
2	236	T/C	H/H	125016	+2	synon	211236	Gallus gallus beta-actin
3	399	A/T	K/I	4865790	+18	nonsyn	2827444	Gallus gallus nucleoside diphosphate kinase
4	516	G/A	T/T	50169290	-4	synon	3914816	Gallus gallus Ribonuclease CL2
5	528	C/T	5' flanking region	8096601	-19		1246528	Gallus gallus gene for acetyl-coA carboxylase
6	806	C/T	Intron	47717691	+1		2198806	Gallus gallus monocarboxylate transporter 3
7	822B	G/C	L/L	143212	-27	synon	222822	Chicken growth hormone gene
8	922	C/T	N/N	67465920	-2	synon	212899	Gallus gallus ovalbumin (ovalY) gene
9	951	G/A	5' flanking region	7137424	-15		14336629	Gallus gallus beta-B1 crystallin gene
10	A9U	A/G	3' untranslated region	62615454	-UN		211814	Chicken retinal glutamine synthetase (Glns)
11	B6U	A/G	G/R	10841687	-6	nonsyn	pgl1c.pk001.b6	microsomal triglyceride transfer protein large subunit [Sus scrofa]
12	C16U	A/G	T/A	10210980	+17	nonsyn	2988353	alpha-1-microglobulin/bikunin [Cavia porcellus]
13	C1U	T/C	5' untranslated region	188033208	+1		16805333	Gallus gallus hemopexin
14	C2	C/T	I/I	12189022	+7	synon	20071147	gene product [Homo sapiens]
15	C22	G/A	UN	UN	UN		pgl1n.pk001.c22	sodium stibogluconate resistance protein [Leishmania tarentolae]
16	D20U	C/T	N/N	4772988	-8	synon	18873598	Gallus gallus flavin-containing monooxygenase 3
17	D3U	T/C	H/H	46456513	-5	synon	pgl1n.pk001.d3	unknown[Best Blast hit- ankyrin repeat domain 9 [Homo sapiens]
18	F11U	T/C	P/P	60464496	-4	synon	2598604	class III alcohol dehydrogenase [Oryctolagus cuniculus]
19	F9U	T/C	D/D	141252700	+UN	synon	pgl1n.pk005.f9	aldose 1-epimerase [Sus scrofa]
20	G16	C/T	UN	9390785	+14		pgl1c.pk001.g16	[None available]
21	H12U	A/G	E/E	99313722	-1	synon	12837772	ATP synthase, H+ transporting [Mus musculus]
22	I18U	A/G	P/P	3637050	-UN	synon	13434994	Gallus gallus PIT 54
23	I4U	C/T	D/D	60504085	-4	synon	62840	Chicken Adh-1 mRNA for class I alcohol dehydrogenase
24	J14U	T/C	5' untranslated region	4747811	+18		211768	Chicken fatty acid synthase gene
25	L13U	G/C	S/S	103087186	+3	synon	pgl1n.pk005.l13	epoxide hydrolase 2 cytoplasmic [Rattus norvegicus]

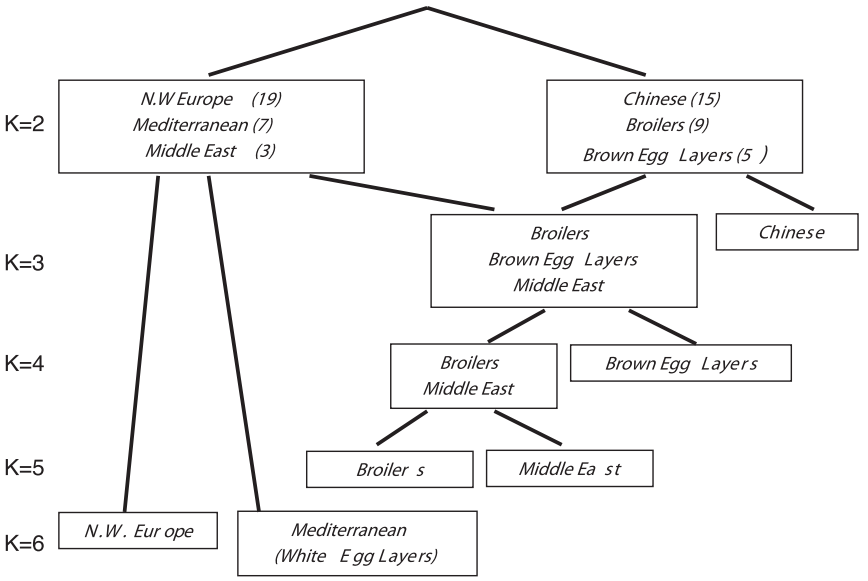
[1] The location of the SNP; in exons, an amino acid change is listed
[2] Genetic map of the SNP in the chicken genome (See M&M),
[a] Position in base pairs.
[b] Chromosome number and strand.
[3] Characterization of the exons' SNPs: synonymous and non-synonymous.
[4] The GI number of NCBI gene bank, or the clone ID of the sequence from EST database (http://www.chickest.udel.edu/).
[5] The name of the gene, or in the case of an EST, the best Blast homology of the chicken sequence. "Gallus gallus" means that the gene was identified in chicken, if not; the best organism in the Blast analysis is given in parenthesis.
Based on Twito *et al.*, 2006. Cytogenetics and Genome Research (in press)

Table 4 Average bootstrap values and their 95% confidence intervals (CI) of seven Neighbour-joining cladograms, varying in marker type, number of birds per population, and number of marker loci.

#	Marker type	Number of birds in each population	Number of loci	Average bootstrap value ± Confidence Interval
1	SSR	30	29	42.4 ± 26.0
2	SSR	10	29	47.4 ± 24.1
3	SNPs in random fragments	10	145*	75.7 ± 17.6
4	SNPs in random fragments	10	14**	29.4 ± 16.0
5	SNPs in genes	10	25	72.7 ± 15.73
6	SNPs in genes	5	25	70.1 ± 16.08
7	SNPs in genes	10	12	42.6 ± 17.6

The cladograms are based on Proportion-Shared Alleles (PSA) using the *microsat* and *PHYLIP* software.
*145 SNPs are distributed across 14 fragments of about 500 bp each; data taken from AVIANDIV project.
**14 SNPs are one SNP from each of the 14 fragments in cladogram #3; data taken from AVIANDIV project.

Figure 1 Schematic structure of six main clusters.



The number of populations in each cluster is presented in brackets.
This schematic tree was constructed according to the most frequent solution among 100 runs; highly similar solutions (C>0.95) repeatedly obtained 100, 24, 11, 9, and 12 times for K=2 to 6, respectively.